

CA2ALED 35

66 T27

CA2 ALED 35 1966T27
Testing In Schools.



3 3398 00132 0737

TESTING IN SCHOOLS



LIBRARY
VAULT 19

SPECIAL SERVICES BRANCH
DEPARTMENT OF EDUCATION
EDMONTON, ALBERTA

1967

1A
1131
A4

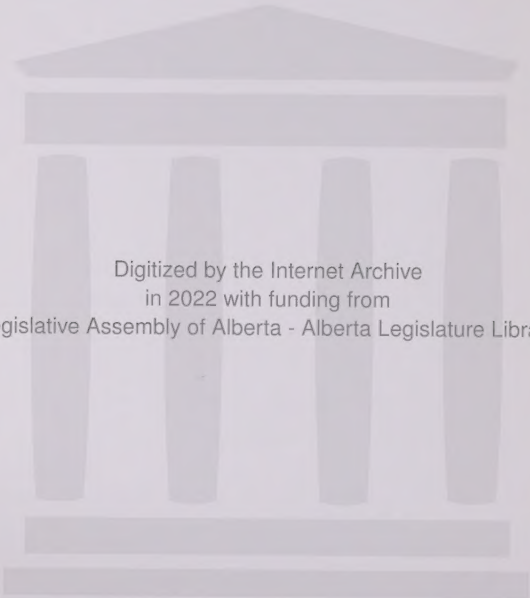
TESTING IN SCHOOLS



HUMAN RESOURCES RESEARCH COUNCIL
11507 - 74 AVENUE
EDMONTON 62, - ALBERTA

SPECIAL SERVICES BRANCH
DEPARTMENT OF EDUCATION
EDMONTON, ALBERTA

1967



Digitized by the Internet Archive
in 2022 with funding from

Legislative Assembly of Alberta - Alberta Legislature Library

FOREWORD

Notes Regarding the Use of this Booklet

Tests are listed in alphabetic order according to the first letters of the first word of the title, e.g. Iowa precedes Metropolitan. Where there are several similar tests of different levels Primary precedes Elementary and Intermediate precedes Advanced. One exception to the system is the general classification "Reading Tests, Instructional" under which three series of tests to accompany primary school readers are listed.

The descriptions that follow are believed to contain much of the information that is essential for the selection of tests. They should be used in conjunction with the Standardized Tests Order Form, issued by the School Book Branch, Department of Education, which is revised every year. If additional information is needed teachers should study specimen sets or consult authorities such as The Sixth Mental Measurements Yearbook, O. K. Buros, (Ed.).

Tests listed in the Order Form that are not described in this booklet are considered obsolete or not too satisfactory.

PART I

DESCRIPTION OF TESTS

A. C. E. PSYCHOLOGICAL EXAMINATION - Grades 10 - 12, 1953.

The American Council on Education Psychological Examination for High School Students is one of the older tests in its field and has proven to be a very reliable measure. The items have been chosen on the basis of results in preliminary trials and weightings obtained by factor analysis. Subtests produce a Linguistic or "L" score, made up of a Similarities - Opposites section and a Completion section, and a Quantitative or "Q" score, consisting of Arithmetic Problems and number series sections. The test booklets are reusable and separate answer sheets must be used. The test is readily scored by means of a perforated scoring key or by machine. The total score, which is more reliable than the subscores gives correlations with college grades that cluster about .50. Subscores are not sufficiently reliable for use in predicting success in scientific subjects and verbal subjects. The test is sold only in packages of twenty-five, each containing a Manual of Instructions, Class Record Sheet, and perforated Scoring Key. The answer sheets are available only in multiples of twenty-five and must be ordered separately. Percentile norms may be obtained on request. Each subtest must be carefully timed; the total time of administration is approximately fifty-four minutes.

CALIFORNIA ACHIEVEMENT TESTS - Grades 1 - 14, 1957.

The California Achievement Tests, revisions of an earlier edition of the series, are designed to measure ability in the basic skills and provide diagnostic scores in several different subjects. The series gives assistance to the teacher in determining educational status of the

pupils and in indicating the particular types of instruction which should be stressed. All levels of the battery produce eleven subscores: vocabulary, reading comprehension, total reading, arithmetic reasoning, arithmetic fundamentals, total arithmetic, English mechanics, spelling, total language, battery total and handwriting. Though the tests are intended to be diagnostic, a student's scores on the subtest should be considered rough measures, only, of his strengths and weaknesses because the subtests are too short to be very reliable. Subscores averaged for a whole class might be considered quite reliable. A profile provided on the cover of the test booklet permits a general diagnosis to be made. The tests are intended to be power rather than speed tests and therefore accurate timing is not as essential to this battery as to some others. Age-grade norms and percentile norms, revised in 1963, are provided for all grade-levels. Local norms would be of value because experience has shown that the publisher's norms are often exceeded by Alberta classes. Each battery is sold in packages of thirty-five tests, including the Manual, Directions for Administering and Scoring Key, and Class Record Sheet. Reading, Arithmetic and Language tests are available separately.

Lower Primary - Grades 1 - 2.

The lower Primary level of the California series is one of the better known achievement batteries intended for the beginning grades in school. Two forms, W and X, are available. The battery requires about two hours to administer, therefore, it must be given in several sections. Booklets are not reusable.

Upper Primary - Grades 3 - 4.5.

Forms W and X of the Upper Primary battery are available, and each requires about two and one-half hours to administer. Booklets are not reusable.

Primary - Grades 1 - 3.

Test booklets are not reusable. About one and one-half hours is required for administering the test.

Elementary - Grades 4 - 8.

Machine-scored answer sheets are available, therefore the booklets are reusable. Administration time is about one and three-quarter hours.

Junior High - Grades 7 - 9.

Test booklets are reusable. Administration time is about two hours.

Secondary - Grades 9 - 13.

Test booklets are reusable. Administration time is about two hours.

Advanced - Grades 10 - 16.

Test booklets are reusable. Administration time is about two hours.

CALIFORNIA SHORT-FORM TEST OF MENTAL MATURITY

- Kindergarten - Adult, 1963.

Eight levels, one form of each, of the 1963 revision of the California Short-Form Test of Mental Maturity are available. The following seven scores are produced: logical reasoning, numerical reasoning, verbal concepts, memory, language total, non-language total and total. The four basic scores in the list were derived by factor analysis. Booklets for Grade 4 and over are reusable; all tests are sold in packages of thirty-five. Tape recorded directions for administration may be purchased. Administration time is much less than for the 1957 tests.

Level 0 - Grades Kindergarten - 1.5.

Administration time is about forty minutes. Booklets are not reusable.

Level 1 - Grades 1 - 3.

Administration time is about forty-five minutes. Booklets are not reusable.

Level 1H - Grades 3 - 4.

Administration time is about fifty minutes. Booklets are not reusable.

Level 2 - Grades 4 - 6.

Administration time is about fifty minutes. Booklets are reusable.

Level 2H - Grades 6 - 7.

Administration time is about forty-five minutes. Booklets are reusable.

Level 3 - Grades 7 - 9.

Administration time is about forty-five minutes. Booklets are reusable.

Level 4 - Grades 9 - 12.

Administration time is about forty-five minutes. Booklets are reusable.

Level 5 - Grades 12 - 16 and adults.

Administration time is about forty-five minutes. Booklets are reusable.

CHICAGO NON-VERBAL EXAMINATION - Age 6 and over, 1947.

This test is used for testing the ability of children who are handicapped in the use of the English language, such as the hard-of-hearing, those with extreme reading difficulties or those from a foreign-language environment. The test should be given individually and may be given in pantomime or with oral directions. (Practice in administration is essential to make sure that the correct "right" or "left" pictures are pointed to immediately.) Subtests must be accurately timed. Scores may be converted to Mental Ages or percentiles, and in combination with Chronological Ages, to scaled scores which are comparable

to deviation I. Q.'s. The test can probably not be used very successfully with pupils below the age of eight. The test is packaged in lots of twenty-five; the Manual and Cutout Scoring Key must be ordered separately. Time of administration is about forty-five minutes.

COOPERATIVE ENGLISH - EFFECTIVENESS OF EXPRESSION, LOWER LEVEL

- Grades 7 - 12, 1943 & 1947.

This test covers aspects of English not included in Mechanics of Expression. It has three parts: I - Sentence Structure and Style (multiple-choice questions based on comparisons of different sentences and paragraphs), II - Active Vocabulary (recall of appropriate words for use in sentences), III - Organization (arranging sentences in order). Three forms, T, X and Y, are available. It is sold in packages of twenty-five reusable test booklets, including the General Manual of Directions, Scoring Key, Class Record, and Percentile Norms. The time of administration is about forty-five minutes.

COOPERATIVE ENGLISH - MECHANICS OF EXPRESSION

- Grades 7 - 16, 1943, 1947, 1950.

This test consists of three parts: I - Grammatical Usage (detection of incorrect forms), II - Punctuation and Capitalization (two to four choices), III - Spelling (two choices). It may be used throughout the junior and senior high school grades. Three forms, T, X and Y, are available. It is sold in packages of twenty-five reusable tests including the Manual of Directions and Interpretation, Scoring Key, and Class Record. The time of administration is about forty-five minutes.

COOPERATIVE ENGLISH C₁ - READING COMPREHENSION, LOWER LEVEL

- Grades 7 - 12, 1943, 1948, 1953.

This test gives three scores: Vocabulary, Speed of Comprehension, and Level of Comprehension. A special perforated scoring key must be ordered if using the answer sheets. Four forms, sold in lots of

twenty-five reusable booklets, are available. The Manual and Key must be ordered separately. The time of administration is about forty-five minutes.

COOPERATIVE FRENCH TEST - 0 - more than two years of High School French, 1941.

This test is a reliable and comprehensive test of vocabulary, reading comprehension, and grammar for advanced French courses. Percentile norms based on U. S. classes are available for different levels depending upon the number of half-years (semesters) that the French language has been studied. Two forms, S and X, are available. Tests are sold only in multiples of twenty-five. The Manual, Key, and Norms are included in the package. Time of administration is about eighty-five minutes.

Lower Level - 1 - 2 years of High School French.

Booklets are reusable.

Higher Level - more than 2 years of High School French.

Booklets are reusable.

COOPERATIVE FRENCH LISTENING COMPREHENSION TEST - 1 - 3 years of High School French, 1955.

This test, available in Forms A and B, may be administered by tape recording or orally. The test consists of four parts: Phonetic Discrimination, Comprehension (reply to questions), Comprehension (sentence completion), Comprehension (reading passage). The 1955 edition is a revision of tests that have been in use for many years.

DETROIT BEGINNING FIRST-GRADE INTELLIGENCE TEST - Grades Kindergarten - 1.

This is a standardized non-reading intelligence test for children, preferably used during the first half of the first grade. It consists of ten pictorial subtests although only a single weighted score is obtained. The weighting is a doubtful feature but must be carried out if comparisons with the norms are to be made. It is doubtful that I. Q.'s

calculated at this level would have much stability, therefore they should not be relied upon. The M. A.'s are useful, however, for the classification of pupils. Only one form (Form A) is available. Tests are sold in packages of thirty-five, including the Manual of Directions and Key, and Class Record Sheet. The test is not timed but requires approximately thirty minutes for administration.

DETROIT ADVANCED FIRST-GRADE INTELLIGENCE TEST - Grade 1 - 2.

This test is useful for classification of children during the second half of the first grade or the first half of the second grade. It may be used for retesting after the Detroit Beginning First-Grade Intelligence Test has been administered. It is a non-reading pictorial test throughout. Seven subtests cover memory, missing parts of pictures, information, similarity of designs, functional parts, counting and ability to read numbers, and classification of objects. Norms are given for chronological age-levels from five and one-half to eleven in terms of letter-grades; and scores may be converted to mental ages. It is doubtful that I. Q.'s that are derived at this level will be of much value, but the mental ages are useful for the classification of pupils. One form only is available. The test is sold in packages of thirty-five including the Manual of Directions and Key, and a Class Record Sheet. Time of administration is thirty to thirty-five minutes.

DIAGNOSTIC READING TESTS - Grades Kindergarten - 13, 1963.

The Diagnostic Reading Tests consist of three batteries of tests, each intended for a different grade range. In each battery there is a survey section intended to measure general proficiency in reading, and a group of supplementary tests for measuring particular skills connected with reading. The survey section, only, of each battery is described here.

Kindergarten Through Fourth Grades, Survey Section - Grades K - 4.

Reading Readiness Booklet - Grades K - 1.

Form B, only, is available which yields five scores: relationships, eye-hand coordination, visual discrimination, auditory discrimination, and vocabulary. Test booklets are not reusable. Some of the testing material is mimeographed and not too satisfactory. Single copies of the booklet are available.

Booklet 1 - Grade 1.

This test yields twelve scores: visual discriminations, auditory discrimination (four scores, including total), vocabulary (four scores, including total), story reading (three scores, including total). Form A only, is available. Single copies of the booklet are sold; they are not reusable. Some material is mimeographed.

Booklet 2 - Grade 2. Booklet 3 - Grades 3 - 4.

These tests yield scores in word recognition and comprehension as well as a total score. Two forms, A and B, have been prepared. Single copies of the booklets, which are reusable, are available. Some materials are mimeographed. Administration time is about thirty minutes for each booklet.

Lower Level, Survey Section - Grades 4 - 8.

Booklet 1: Part 1, Word Recognition and Comprehension.

This test, available in forms A, B, C and D, yields two scores as listed in the title. The test booklets, sold as individual copies, are reusable. Administration time is about thirty minutes.

Booklet 2: Part 2 and Part 3.

This test yields three scores: vocabulary, speed of reading and story comprehension. Four forms, A - D, are available. Test booklets, sold individually, are reusable. Administration time is about thirty minutes.

Upper Level, Survey Section - Grades 7 - 13.

The Diagnostic Reading Test, Upper Level: Survey Section produces five scores: rate of reading, comprehension, vocabulary, total comprehension, and total. Eight forms, A - H, of this test are available; they are dated from 1947 to 1952. Test booklets, obtainable individually, are reusable. Administration time is slightly less than one hour.

DIFFERENTIAL APTITUDE TESTS - Grades 8 - 12 and adults, 1963.

The Differential Aptitude Tests (D.A.T.) are becoming increasingly

popular among high school guidance counsellors for providing information useful for advising students as to what occupational choices they should consider. The battery is sometimes employed after an interest test has been given; in such instances it is scarcely necessary to give the full D.A.T. battery because areas in which a student has shown a low interest need not be explored further. The battery yields nine scores: verbal reasoning, numerical ability, scholastic aptitude (total), abstract reasoning, space relations, mechanical reasoning, clerical speed and accuracy, spelling, and sentences. Two new forms, L and M, are available. Both forms employ Optical Scoring Reader (IBM 1230) Answer Sheets. Forms L and M are each contained in two booklets. When ordering answer sheets, it is important that you state the Form (L or M) required since the answer sheets are not applicable to either Form. Administration time is about two hours for each booklet. Up to six sessions may be used when the tests are given.

GATES-MACGINITIE READING TESTS - Grades K - 9.

The Gates-MacGinitie Reading Tests are a new (1965) series replacing the 1958 revision of the Gates Reading Tests. The series consists of two or three forms at each of six levels. An additional group, intended for high school use, are not yet ready for purchase. At levels involving grade four and higher, the tests have been adapted for machine scoring.

Gates-MacGinitie Primary A Reading Tests - Grade 1.

This test produces a vocabulary score and a comprehension score. Two forms, 1 and 2, are available in packages of thirty-five. Total testing time is about forty minutes. Booklets are not reusable

Gates-MacGinitie Primary B Reading Tests - Grade 2.

Details for these tests are the same as for the Primary A series except that they are intended for use in Grade 2.

Gates-MacGinitie Primary C Reading Tests - Grade 3

Details for these tests are the same as for the Primary A series except that they are intended for Grade 3, and the testing time is about fifty minutes.

Gates-MacGinitie Primary CS Reading Tests - Grades 2 - 3.

The Primary CS test measures speed of reading only. It is available in three forms, 1, 2 and 3, and packaged in groups of thirty-five booklets; these booklets are not reusable. Testing time is seven minutes.

Gates-MacGinitie Survey D Reading Tests - Grades 4 - 6.

The Survey D tests yield a speed and accuracy score, a vocabulary score, and a comprehension score. Forms 1M, 2M and 3M, each sold in packages of thirty-five are available. These forms have been adapted for IBM 1230 machine scoring, therefore the booklets are reusable. Total testing time is forty-five minutes.

Gates-MacGinitie Survey E Reading Tests - Grades 7 - 9.

The description of the Survey D section applies to the Survey E section except for the grade level.

GATES READING SURVEY - Grades 4 - 10.

The Gates Reading Survey may be used with average students from the end of Grade 4 to the lower grades of high school, but is most suitable from Grades 5 to 7. It is especially useful for general survey purposes and may be used to select pupils who require diagnosis as provided by the Basic Reading Tests. The four fundamental aspects of reading ability that are measured by the Reading Survey are: Vocabulary (the words have been chosen from the Thorndike list of words most commonly used in representative reading material), Level of Comprehension, Speed and Accuracy, and Accuracy. Directions to the examiner are given inside the front cover of the test booklet. The Manual contains additional instructions for scoring and for the interpretation of test results. Three forms, M1, M2 and M3, are available. The tests are sold in packages of thirty-five. The time of administration is about seventy minutes.

GENERAL CLERICAL TEST (PSYCHOLOGICAL CORPORATION) - Grades 9 - 12, 1950.

The Psychological Corporation's General Clerical Test is a combined aptitude and achievement test to distinguish between abilities in different types of clerical tasks. It contains nine subtests. The over-all score may be regarded as an index of general clerical aptitude. To a certain extent, the nine sub-scores may be used to help guide students toward special clerical jobs. The nine subtests are: Checking Copy, Alphabetization, Arithmetic Computation, Checking Addition, Arithmetic Problems, Spelling, Reading Comprehension, Vocabulary, and Correction of Grammatical Errors. Norms are given in terms of percentiles for high-school girls, office employees, and industrial workers for Clerical Speed and Accuracy, Numerical Ability and Verbal Facility. Only one form is available. Tests are sold in packages of twenty-five, including Manual and Scoring Keys. The time of administration is approximately fifty minutes.

GINN BASIC READING TESTS: - (See READING TESTS, INSTRUCTIONAL) Page 21

GRAY ORAL READING TEST - Grades 1 - 12 and adult, 1963.

The Gray Oral Reading Test is one of the newer tests in its field and must be considered one of the better ones. It is an individual test that requires use of a reading passage booklet that is reusable and a record booklet that is not reusable. Four forms of the test (A - D) are available; individual copies of the reading passage booklet are sold and the record booklets are packaged in units of thirty-five, with a manual included. Norms are tentative, therefore they might not be too satisfactory at present.

IOWA TESTS OF BASIC SKILLS - Grades 3 - 9, 1955.

One of the best achievement batteries available is the Iowa Tests of Basic Skills. It appears to be expensive, since the cost per test

booklet is high relative to other tests, but the price seems reasonable when it is borne in mind that it can be used for seven grades and yields the following fifteen scores: Vocabulary, Reading Comprehension, five Language scores, four Work-Study Skills scores, three Arithmetic Skills scores and a Total score. A few of the items have a culture-bias in that some of the reading passages deal with American history, and American coins are pictured in the arithmetic test, but the effects of this should not be serious. Three forms, 1, 2 and 3, are now available. Grade norms and percentile norms within each grade are available. Norms for some Alberta systems are also available upon request. Administration time, in four sessions, is about five and one-half hours

A Revised Canadian Edition called The Canadian Tests of Basic Skills (Grades 3 to 8) is also available in two forms, Form 1 and Form 2. Canadian norms can be obtained. At the present time only MRC answer sheets (hand-scoring) are available. IBM 1230 answer sheets should be available in the very near future. The Canadian author is Dr. Ethel King of the University of Calgary.

IOWA TESTS OF EDUCATIONAL DEVELOPMENT - Grades 9 - 12.

The Iowa Tests of Educational Development is a battery of nine tests measuring four areas of curriculum. Ten scores are produced, they are: understanding of basic social concepts, general background in the natural sciences, correctness and appropriateness of expression, ability to do quantitative thinking, ability to interpret reading materials in the social studies, ability to interpret reading material in the natural sciences, ability to interpret literary materials, general vocabulary, subtotal, and use of sources of information. The tests are widely used in the United States but some of the tests are not particularly suitable for use in Alberta. Since the tests are sold as

separate booklets, it is possible to avoid purchasing material that will not be useful. Two forms, X-3S and Y-3S, both adapted for machine scoring, are available. Time allowances for the tests are rather liberal because they are intended to be power tests. However, in order to make the tests fit into regular class periods, shorter time limits may be used. Testing time, for the full battery, is about five hours if the shorter times are used. Tests are sold in packages of twenty.

KUDER PREFERENCE RECORD- Vocational - Grades 9 - 12, 1962.

The Kuder Preference Record - Vocational is designed to measure fundamental areas of educational and vocational interest. It is not an aptitude or ability test. It yields eleven scores: Mechanical, Computational, Scientific, Persuasive, Artistic, Literary, Musical, Social Service, Clerical, Outdoor Activities, Masculinity-Femininity, and a verification score. Form CH of the test is arranged in loose-leaf booklets with thirteen step-down pages. A self-scoring answer pad is to be inserted in the booklet and answers are indicated by perforation with a pin. Opening the answer pad permits counting of the perforations and consequently scoring of the ten areas of interest. Scores for different areas are plotted on a profile sheet from which they may be converted directly to percentiles. Interest patterns determined from the profile may be related to different types of occupations that are classified in the manual. If judiciously used by qualified persons the Preference Record provides an excellent basis for discussion, and providing that the pupil's interests have satisfactorily crystallized, it indicates the type of vocational field

in which the pupil might be wise to seek employment. It is not, of course, a measure of ability in any vocational field. Several forms of the test are available; the most popular are Forms CH and CM, the hand-scored and machine-scored versions, respectively, of the 1948 edition with manuals revised in 1962. The inventory is untimed but may be completed and scored in forty to fifty minutes.

KUHLMANN-ANDERSON INTELLIGENCE TESTS - Grades K - 12, 1963.

The seventh edition of the Kuhlmann-Anderson Intelligence Tests is a series of tests at eight levels: Grades K, 1, 2, 3 - 4, 4 - 5, 5 - 7, 7 - 9, and 9 - 12 labelled, respectively, booklets K, A, B, CD, D, EF, G and H. One form of each test, only, is available. Booklets K, A, B, CD, D and EF (Grades K - 7) produce a total score which can be converted to percentile ranks by grade, or to deviation I. Q.'s. The remaining tests for Grades 7 - 12, booklets G and H, yield V, Q and total scores, each of which can be converted to grade percentiles. Separate administration and norms manuals are required at each level, though one technical manual serves for all levels. Separate answer sheets are available for levels D - H. Administration time for Booklets K, A, B, and CD is about forty-five minutes, for the remainder it is about one hour each. Tests are sold in packages of twenty-five. This test is considered to be among the best available for measuring intelligence.

LORGE-THORNDIKE INTELLIGENCE TESTS - Grades K - 12.

There are five separate levels of the Lorge-Thorndike Intelligence Tests with two forms at each level. Tests for Grades K - 3 are non-verbal, and for Grades 4 - 12 there are both verbal and non-verbal

tests available, though it is intended that both tests be given to each student. Both the verbal and non-verbal tests for Grades 4 - 12 may be machine-scored, therefore the test booklets are reusable. This series of tests is among the best available for measuring intelligence.

Level 1 - Grades K - 1.

This non-verbal test measures comprehension of oral language and abstract thinking. The two forms, A and B, are sold in packages of thirty-five. Testing time is thirty-five minutes in three sittings. Booklets are not reusable.

Level 2 - Grades 2 - 3.

The description is the same as for Level 1.

Level 3, Non-Verbal - Grades 4 - 6.

This test is intended to measure spatial reasoning and it is suggested that it be used with the Level 3, Verbal test. Forms A and B, sold in packages of thirty-five reusable booklets, are available. Testing time is thirty minutes.

Level 3, Verbal - Grades 4 - 6.

This test is designed to measure the abilities associated with scholastic aptitude. The reusable test booklets, sold in packages of thirty-five, are available in Forms A and B. Testing time is thirty-minutes.

Level 4, Non-Verbal - Grades 7 - 9.

Though only one score, abstract reasoning ability, is produced, there are three subtests, as follows: pictorial classification, pictorial analogy and numerical relationships. Forms A and B, in packages of thirty-five reusable test booklets, are available. Testing time is about thirty minutes.

Level 4, Verbal - Grades 7 - 9.

This test has subtests in vocabulary, verbal classification, sentence completion, arithmetic reasoning and verbal analogy, but subscores are not produced. Forms A and B, in packages of thirty-five reusable booklets, are available. Testing time is about thirty-five minutes.

Level 5, Non-Verbal - Grades 10 - 12.

The description is the same as for Level 4, Non-Verbal.

Level 5, Verbal - Grades 10 - 12.

The description is the same as for Level 4, Verbal.

METROPOLITAN ACHIEVEMENT TESTS - Grades 1 - 12, 1958, 1963.

The Metropolitan Achievement Tests are a coordinated series of eight test batteries covering major subject matter fields in steps from Grades 1 to 12. The Primary I and Primary II batteries are intended for the corresponding single grades but may be used in the earlier portion of the succeeding grades. The Elementary, Intermediate, and Advanced tests each cover two grades and may be used in the early months of the succeeding grades. The Intermediate Complete and Advanced Complete tests are not advocated for use in Alberta because the complete batteries at these levels contain a test based largely on American history and other tests that are not suitable for the Alberta curriculum; for that reason they are not described below. Batteries producing a social studies score have a few items based on American history that might be omitted. Each test booklet has age and grade equivalents and a profile chart printed on the cover. Those using the profile should notice that the range in terms of grade-equivalents is much less for some subtests than for others. That results in a characteristic profile from which it would appear that all bright pupils are relatively poor and all dull pupils are relatively good in arithmetic fundamentals. Each package of thirty-five tests includes a Class Analysis Chart, Scoring Keys, Directions for Administration, and a Class Record Sheet. The Manual for Interpreting the Metropolitan Achievement Tests is a 122-page publication that contains valuable information in regard to testing in general, and suggestions for the use of the test scores. A separate booklet of United States percentile norms may be specially

ordered; however, schools are advised to construct their own norms.

Primary I Battery - Grade 1.

The Metropolitan Primary I Battery should not be administered until late in the first school year. A single booklet contains tests in Reading - Word Picture Identification, Word Recognition and Meaning, and Numbers. Forms A, B and C are available. The tests are sold in packages of thirty-five, containing Manual of Directions, Key, Norms, Class Record and Class Analysis Sheet. Administration time is about two hours. Testing should be done in four sessions.

Primary II Battery - Grade 2.

A single booklet contains the tests that yield scores in Word Knowledge, Word Discrimination, Reading, Spelling, and three Arithmetic scores. Administration time is about two hours. The tests should be given in four sessions. Forms A, B and C are available.

Elementary Battery - Grades 3 - 4.

The test provides nine scores: Word Knowledge, Word Discrimination, Reading, Spelling, three Language scores and two Arithmetic scores. Forms A, B, C and D are available. Administration time is about three hours, spread over five sessions.

Intermediate Partial Battery - Grades 5 - 6.

The test produces eleven scores: Word Knowledge, Reading, Spelling, Language Usage, Punctuation and Capitalization, Total Language, Language Study Skills, Computation, Problem Solving and Concepts, and Social Studies Skills. Forms A, B, C and D of a hand-scored edition are available, as well as forms AM, BM, CM and DM of a machine-scored edition. IBM 1230 answer sheets are available only for Forms AM and BM. The subtests are fairly reliable, except perhaps for the problem solving and concepts test. Administration time is slightly less than four hours; five sessions are recommended.

Advanced Partial Battery - Grades 7 - 9.

The tests yield the same eleven scores as the Intermediate Partial Battery plus an additional language score called Kinds of Sentences. The same eight forms as for the Intermediate Partial Battery are available. Administration time is about four hours; five sessions are recommended.

High School Battery - Grades 9 - 12.

The following eleven scores are produced: Reading, Spelling, Language, Language Study Skills, three Social Studies scores, two

Mathematics scores, and two Science scores. Two forms, AM and BM, both requiring separate answer sheets, are available. Administration time for the battery is about five and one quarter hours, but this might be shortened by omitting some of the tests that are intended for American students. Local norms should be developed and used in preference to the American norms.

METROPOLITAN READINESS TEST - Grades K - 1, 1964.

This is a test for children of beginning school age which measures the maturity of certain mental functions that are involved in the learning of first-grade work. Maturity also indicates the pupil's level of ability to do abstract thinking, but it is not desirable to convert scores to I.Q.'s at such an early age. The tests are valuable for purposes of grouping first-grade pupils for instruction although later re-grouping undoubtedly will be required. The probability of success increases with increase in readiness. The subtests are completely pictorial and cover Word Meaning, Sentences, Information, Matching, Numbers and Copying (visual perception and motor control). Three scores are obtained - Reading Readiness (tests 1 to 4), Number Readiness (test 5), and Total Readiness (the total score). Results will indicate the pupils who may safely be taught first-grade work and those who will need certain special training before "academic" work is begun. Total Readiness will give higher correlations with achievement than will either of the subscores. The correlation that may be expected between Total Readiness scores and achievement at the end of the first year is approximately .5 with average reading and .6 with number work. Two forms, A and B, are available. The test is sold in packages of thirty-five, including the Manual, Key, and Class Record. The total administration time of approximately seventy-five minutes should be broken up into at least three sittings.

MINNESOTA CLERICAL TEST - Grades 8 - 12 and adult.

The test was devised in 1933 but a manual revised in 1959 is now available. Two scores are produced - Number Comparison and Name Comparison. Though the test is old it is considered one of the best in its field. Administration time is twenty minutes. The test, available in one form only, does not have reusable booklets.

OTIS QUICK-SCORING MENTAL ABILITY TEST - ALPHA SHORT-FORM

- Grades 1 - 4, 1954.

Two tests can be given with the same test papers and may be scored by means of the same key. (It is preferable that pupils use pencils of different colours.) There are two sets of directions, each referring to the same set of pictures. This results in two scores: non-verbal and verbal. The verbal directions are much more detailed and complete and involve greater experience with the English language. The non-verbal test is really a non-reading test. One form of the test is available. Tests are sold in packages of thirty-five, including the Manual of Directions, cut-out Scoring Key, and a Class Record Sheet. Administration time is about thirty-five minutes.

OTIS QUICK-SCORING MENTAL ABILITY TEST - BETA - Grades 4 - 9.

The Otis Beta Test is self-administering. It contains eighty items, of which two-thirds involve verbal competence. The booklets are so constructed that the cover may be used as the answer sheet, which is then scored by means of a perforated scoring key. If reuse of the tests is desirable, a separate answer sheet may be ordered as a substitute for the cover. Forms Em and Fm are available, as well as some earlier forms. The tests are sold in packages of thirty-five, including the Manual of Directions, Scoring Key, and Class Record. Answer sheets are optional and may be ordered in multiples of ten. Administration time is about thirty-five minutes.

OTIS QUICK-SCORING MENTAL ABILITY TEST - GAMMA - Grades 9 - 16.

The Otis Gamma Test is the highest in the Quick-Scoring Mental Ability Test series. It is useful throughout the senior high school grades. It contains eighty items which are weighted heavily for verbal ability. The booklet is constructed so that the cover may be used as a separate answer sheet, which is readily scored by means of a perforated scoring key. Separate answer sheets may be used if it is desirable to keep the test booklets for reuse. A single I.Q. score is produced. Forms Em and Fm, as well as some earlier forms, are available. The tests are sold in packages of thirty-five, including the Manual, Scoring Key, and Class Record. Administration time is about thirty-five minutes.

READING TESTS, INSTRUCTIONAL

For each series of readers authorized for elementary grades in Alberta there is available a parallel series of tests. For each reader a test has been prepared to be administered when a class or a pupil has the work connected with the reader. In some cases a pre-test is also available to be administered before a particular reader is studied. The tests are termed "instructional" because their primary purpose is to determine the effectiveness of the instruction in reading as suggested by the publisher when a particular series of readers is used. The content of the tests is connected rather closely to the material presented in a particular series of readers. The tests were not designed to measure reading in the sense of an ordinary standardized test where the teaching methods used should be incidental. The tests, then, can assist the teacher in deciding what instructional techniques to use, how pupils should be grouped, what remedial devices are necessary,

and in general, how each class is developing during the school term.

Three series of tests for primary grades are available as follows:

- (a) Basic Reading Tests to accompany the New Curriculum Foundation Series.
- (b) Basic Reading Tests to accompany the Reading for Meaning Series.
- (c) Ginn Basic Readers Achievement Tests to accompany Ginn Basic Readers.

Additional tests are available for grades above primary school, and additional tests become available from time to time as they are prepared. Details regarding the tests are not given here. The Standardized Tests Order Form, issued by the School Book Branch, Department of Education, states which test accompanies each reader.

SASKATOON PUBLIC SCHOOLS HANDWRITING SCALES

The handwriting scales for Grades 1 to 8 have been prepared by the Saskatoon Public Schools for use in evaluating handwriting. Samples of handwriting are given for comparison purposes.

STANFORD ACHIEVEMENT TESTS - Grades 1 - 9, 1964.

The Stanford Achievement Test Battery is one of the better series among the large number of achievement tests that may be purchased, especially if content coverage is required. Scores in the various tests may be transformed into grade equivalents, percentile ranks and stanines. In addition, percentile ranks are available for three different dates of testing, beginning, middle and end of year. Norms are reasonably satisfactory for Alberta. The tests in spelling and language, arithmetic, reading and the intermediate battery tests in science and social studies are available as separate booklets. Separate answer sheets are available for tests involving Grade 4 and above.

Though two or three forms of each test are now on the market, the publisher plans to produce a total of four at each level.

Primary 1 Battery - Grades 1.2 - 2.5.

Six scores are produced: Word Reading, Paragraph Meaning, Vocabulary, Spelling, Word Study Skills, and Arithmetic. Forms W, X and Y, are available; administration time is about two and one-half hours, with five sessions recommended.

Primary 2 Battery - Grades 2.5 - 3.9.

Eight scores are produced: Word Meaning, Paragraph Meaning, Science and Social Studies Concepts, Spelling, Word Study Skills, Language, Arithmetic Computation, and Arithmetic Concepts. Forms W, X and Y are available; administration time, over seven recommended sessions, is about four hours.

Intermediate 1 Battery (Partial Battery) - Grades 4 - 5.5.

Eight scores are produced as follows: Word Meaning, Paragraph Meaning, Spelling, Word Study Skills, Language, Arithmetic Computation, Arithmetic Concepts, and Arithmetic Applications. The complete battery, not described here, produces two additional scores, Social Studies and Science. Forms W, X and Y are available; testing time is about four hours with five sessions recommended.

Intermediate 2 Battery (Partial Battery) - Grades 5.5 - 6.9.

Seven subscores are produced, the same as those produced by the Intermediate 1 series except that Word Study Skills is omitted. This applies to the complete battery also. Administration time is about three and three-quarter hours, in five sessions. Forms W, X and Y are available.

Advanced Battery (Partial Battery) - Grades 7 - 9.

The advanced battery yields six scores: Paragraph Meaning, Spelling, Language, Arithmetic Computation, Arithmetic Concepts, and Arithmetic Applications. Forms W, X and Y are available. Testing time, in four sessions, is a little less than three and one half hours.

High School Battery - Grades 9 - 12.

The high school battery produces seven subscores as follows: English, Numerical Competence, Mathematics, Reading, Science, Social Studies and Spelling. Supplementary tests in Arts and Humanities, Business and Economics, and Technical Comprehension have been prepared. Two forms, W and X, of the basic tests are available, but only one

form, X, of each of the supplementary tests is on the market.
Testing time, in six sittings, is about five and one-half hours.

PART II

The purpose of PART II of this booklet is to provide some basic information to teachers and administrators charged with the responsibility of testing in Alberta schools. The information is given under three headings: Common Measurement Terms, Test Norms, and Some Valuable Techniques.

COMMON MEASUREMENT TERMS

Battery. A series of tests designed to cover a certain field, and intended to be administered, in order, to a group of people. Most batteries employed in schools are series of achievement tests covering the basic school subjects. Comparable scores on each subtest are usually produced.

Aptitude Test. A test designed to assess the potential of a person rather than how much he has already learned. Many aptitude tests measure in very restricted areas such as musical aptitude or clerical aptitude, but some are intended to be more general.

Achievement Test. A test to determine how well a student has grasped what he has been taught or how proficient he is in certain skills. Most tests given by teachers are achievement tests.

Content Validity. The degree to which a test has sampled the content of the area that is to be measured, and the degree to which the proportions of items selected from each is fair.

Evaluation. Determining the value, worth, or merit of something. In education, it involves deciding to what extent the objectives of education have been attained. Evaluation involves more than measurement.

Factor Analysis. A complex mathematical technique used to explain

correlations between many pairs of measures in terms of a few basic characteristics or factors.

Grade Norm. The average score achieved by pupils at a certain grade-level; e.g., if a score of thirty-four corresponds to a grade norm of 6.4 this means that the average score of pupils in the fourth month of Grade 6 is thirty-four.

Average. A single number used to represent a whole group of scores. There are several "averages" in common use, such as the mean, median and mode.

Age Norm. The average score achieved by people of a certain age. If a score of twenty-seven corresponds to an age norm of 8.11, this means that twenty-seven is the average score achieved by people eight years and eleven months old.

Measurement. In education, the function of measurement is to assign numbers to people on the basis of some characteristic that they possess. It is not too difficult to represent achievement, intelligence and even attitude in terms of numbers.

Percentile Rank. The percentile rank of a score is a number indicating what percentage of all the scores fall below the given score. For example, if a score of 129 has a percentile rank of sixty-two, that means that sixty-two percent of the scores were less than 129.

Power Test. A test in which time is not an important factor; i.e., if students were given more time they would not change their scores to any appreciable extent.

Speed Test. A test involving a simple task but when the speed of performing this task is to be measured.

Reliability. A measure of how consistent the scores on a test are.

Usually, a reliability coefficient of a test is given. This is a correlation between scores on a test and an equivalent form, or between scores on two administrations of a test to the same people. There are also ways of determining a reliability coefficient from one administration of a test.

Test. Any device which compares, systematically, the behavior of two or more people. Tests are usually classified according to the psychological characteristics that they measure, such as achievement, aptitude, interest, personality, etc.; however, they may be classified by technical features, such as individual vs. group, speed vs. power, etc..

Diagnostic Test. A test whose function is to locate specific strengths and weaknesses of pupils, as opposed to yielding a single, over-all score.

ORGANIZING A TESTING PROGRAM

Many schools have a regular schedule of standardized tests that are given every year. Such a schedule, or testing program, can be of value in providing information for improving instruction, classifying pupils and evaluating achievement. The value of a testing program is not determined by its elaborateness, but by the extent to which results are put to use.

To encourage the use of test results, decisions regarding what tests to include in a program should be made during discussions involving the entire school staff. Before any test is adopted for use it should be clearly understood how the results will be used.

In deciding what tests to use a booklet such as this, which describes some tests, should be consulted. It is wise, however, to study

specimen sets of tests before they are ordered in quantity. The most important questions to be answered in selecting a test are as follows:

- a) Is the test valid? This is best determined by checking on what subscores are produced and by studying the items to see if they cover the content required.
- b) Is the test reliable? Test reliability is best determined by consulting the manuals. A reliability coefficient of .9 is usually adequate, but not if the test is to be used to make important decisions regarding individuals.
- c) Are the norms applicable? The basic matter here is whether the publisher's norms were established on the basis of a group comparable to the group to be tested. If the norms are not applicable the test can still be used provided local norms are available.
- d) Is the test efficient? Factors such as how easy the test is to administer, score, and interpret are included in the efficiency of a test. This can be judged by consulting the manual and the test itself.

A testing program should be so organized that every member of the staff has a responsibility; also, every member of the staff should have an interest in some aspect of the results.

When a program has been completed, results should be entered on pupil records as soon as possible. It is important that the records be readily available to staff members when required.

SOME WORDS OF CAUTION

- a) All educational measurements are subject to error, therefore, small differences between pupil's scores should not be regarded as meaningful. If there are wide deviations between test results and expectation it is wise to retest, using a different form of the test.
- b) While determining I.Q.'s for primary pupils has some value, the disadvantages often outweigh the advantages. Results of group intelligence tests tend to be unreliable, therefore, there is a danger that pupils might be misclassified and suffer for it. If there is a danger that a teacher will misunderstand the tentative nature of I.Q.'s at the primary level it is just as well that the I.Q.'s are not determined.

- c) When standardized tests are being administered it is important that the instructions be followed to the letter.
- d) Established testing programs should be reviewed every few years; tests go out of date and purposes for testing change.
- e) Whenever possible, separate answer sheets should be used so that tests can be scored by machine or by clerks. A teacher's time is better spent on deciding upon courses of action based on test results.
- f) Test norms are of value only to the extent that the group of pupils on which the norms are based resemble the group of pupils to be tested. It sometimes happens that a good test does not have norms applicable to a certain group to be tested. In cases of this kind local norms might be constructed.
- g) When important decisions regarding pupils must be made, such as "streaming", and promotion or non-promotion, it is essential that reliable information, from tests and other sources be available. Though many things affect reliability, one of the most important factors is test length. It is not wise, therefore, to make important decisions on the basis of a single measure, especially if only an hour or so of testing is involved. Similarly, decisions based on subjective judgments, which are notoriously unreliable, are probably not the wisest ones.

PART III

MULTIPLE-CHOICE VS. ESSAY TEST

Teachers frequently inquire as to the effectiveness of multiple-choice tests as compared to essay tests. The advantages and disadvantages of each are outlined below:

Advantages of Essay Tests

1. They are relatively easy to construct.
2. They can be used to measure how well a student can organize his work and how he can express himself.
3. A student's knowledge can be probed in depth with relatively few questions.

Disadvantages of Essay Tests

1. Essay Tests are tedious to score.
2. They are very difficult to score accurately; i.e., an independent grading of a set of papers will yield wide discrepancies.
3. Speed of writing may be a factor.
4. Bluffing, on the part of the student, is often a problem.

Advantages of Multiple-Choice Tests

1. They are easily scored with great accuracy.
2. There are mathematical procedures for judging the quality of a test and of test questions, making it possible to improve particular questions on the basis of previous experience.
3. Multiple-choice questions lend themselves to repeated use.
4. A wide content area can be covered in a relatively short time.
5. An examiner can control the difficulty level and the content emphasis of a test much easier if multiple-choice questions are used.

Disadvantages of Multiple-Choice Tests

1. Guessing is a problem in multiple-choice tests.
2. The value of a multiple-choice test depends a great deal on the skill of the test constructor.

3. Multiple-choice tests are laborious to construct.

4. Speed of reading may be a factor.

A matter of special concern to teachers is the effect of guessing in a multiple-choice test; in particular they would like to know how guessing may affect students scores. It is not possible to give an answer in simple terms that covers all possibilities, therefore, some cases will be described to serve as examples.

Let us suppose that a test consisting of 100 multiple-choice questions, each with four suggested answers, is administered to a large number of students. Let us also assume that we have precise knowledge as to how many of the answers each of the students actually knew and on how many items he was forced to guess.

TABLE I, below, shows the range of scores that would occur for ten groups of 1,000 students. It is assumed that each group of 1,000 knew the answers to a fixed number of questions and guessed at the remaining answers. It is assumed further that when students responded to items for which they did not know the answers, they responded entirely on the basis of chance.

TABLE I

Number of answers known	Number of answers attempted by guessing	Mean Score not corrected for guessing	Range of middle 800 scores, not corrected for guessing	Mean Score Corrected for guessing	Range of Middle 800 scores, corrected for guessing
0	100	25.0	19 - 31	0	-8 - 8
10	90	32.5	27 - 38	10	3 - 18
20	80	40.0	35 - 45	20	14 - 27
30	70	47.5	43 - 52	30	24 - 36
40	60	55.0	51 - 59	40	35 - 46
50	50	62.5	59 - 66	50	46 - 55
60	40	70.0	66 - 74	60	55 - 66
70	30	77.5	74 - 81	70	66 - 75
80	20	85.0	83 - 87	80	78 - 83
90	10	92.5	91 - 94	90	88 - 92

In order to classify the information given in TABLE I the fifth row of figures is interpreted below.

If 40 answers were actually known by a group of 1,000 students and answers for the remaining 60 were actually guessed, the mean score of the group would be 55. The range of scores of the middle 800 would be 51 - 59. If the scores were corrected for guessing, using the standard formula $R - \frac{W}{3}$, the mean score would be 40 and the range of the middle 800 would be 35 - 46.

It was stated previously that the score ranges given in TABLE I are based on a hypothetical test consisting of 100 items, each with four suggested answers. The range of scores relative to the total score would be smaller if the number of items were increased. There would be a marked reduction in the range if the number of choices were increased.

LB Alberta. Dept. of Education.
1131 Special Services Branch.
A4 Testing in schools. Edmon-
ton, 1966.

